

En este proyecto aplicarás los conocimientos aprendidos hasta el momento. Deberás realizarlo **una vez hayas completado todo el contenido del módulo "Python for data"**.

### **Objetivo del proyecto.**

Realizar un análisis exploratorio de los datos que puedes encontrar que te aportamos. Para realizar este análisis es obligatorio realizarlo con Python.

### **Requisitos del proyecto.**

A lo largo del proyecto tienes que cubrir los siguientes puntos:

- Transformación y limpieza de los datos.
- Análisis descriptivo de los datos.
- Visualización de los datos.
- Informe explicativo del análisis.

### **Herramientas para realizar el proyecto.**

Tendrás que realizar el proyecto usando estas herramientas:

- Python
- Pandas
- Visual Studio Code

Los datos con los que vas a trabajar son los que te aportamos en este [link](#).

Estos conjuntos de datos están relacionados con campañas de *marketing* directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente para determinar si el producto (depósito a plazo bancario) sería suscrito o no. Las columnas que tenemos en el primer *dataset* ('**bank-additional.csv**') son:

- **age**: La edad del cliente.
- **job**: La ocupación o profesión del cliente.
- **marital**: El estado civil del cliente.
- **education**: El nivel educativo del cliente.
- **default**: Indica si el cliente tiene algún historial de incumplimiento de pagos (1: Sí, 0: No).
- **housing**: Indica si el cliente tiene un préstamo hipotecario (1: Sí, 0: No).
- **loan**: Indica si el cliente tiene algún otro tipo de préstamo (1: Sí, 0: No).
- **contact**: El método de contacto utilizado para comunicarse con el cliente.
- **duration**: La duración en segundos de la última interacción con el cliente.
- **campaign**: El número de contactos realizados durante esta campaña para este cliente.
- **pdays**: Número de días que han pasado desde la última vez que se contactó con el cliente durante esta campaña.
- **previous**: Número de veces que se ha contactado con el cliente antes de esta campaña.
- **poutcome**: Resultado de la campaña de marketing anterior.
- **emp.var.rate**: La tasa de variación del empleo.
- **cons.price.idx**: El índice de precios al consumidor.
- **cons.conf.idx**: El índice de confianza del consumidor.
- **euribor3m**: La tasa de interés de referencia a tres meses.

- **nr.employed**: El número de empleados.
- **y**: Indica si el cliente ha suscrito un producto o servicio (Sí/No).
- **date**: La fecha en la que se realizó la interacción con el cliente.
- **contact\_month**: Mes en el que se realizó la interacción con el cliente durante la campaña de marketing.
- **contact\_year**: Año en el que se realizó la interacción con el cliente durante la campaña de marketing.
- **id\_**: Un identificador único para cada registro en el dataset.

El segundo set de datos ('**customer-details.xlsx**') es un archivo Excel que nos da información sobre las características demográficas y comportamiento de compra de los clientes del banco. Este Excel consta de 3 hojas de trabajo diferentes, en cada una de ellas tenemos los clientes que entraron en el banco en diferentes años. Sus columnas son:

- **Income**: Representa el ingreso anual del cliente en términos monetarios.
- **Kidhome**: Indica el número de niños en el hogar del cliente.
- **Teenhome**: Indica el número de adolescentes en el hogar del cliente.
- **Dt\_Customer**: Representa la fecha en que el cliente se convirtió en cliente de la empresa.
- **NumWebVisitsMonth**: Indica la cantidad de visitas mensuales del cliente al sitio web de la empresa.
- **ID**: Identificador único del cliente.

### **Método de entrega.**

La entrega del proyecto se hará a través de GitHub. Para entregarlo tendrás que copiar la url del repositorio de GitHub, asegurate que el repositorio está publicado como público hasta la finalización del curso, requisito indispensable para poder realizar la corrección. Tu repositorio tiene que constar, al menos, de los siguientes archivos/carpetas:

- Archivo README.md, que recoja los pasos seguidos durante el proyecto y el informe de tú análisis.
- Una carpeta de datos donde guardes los archivos en bruto, asociados a este proyecto, y los datos guardados después de las transformaciones.
- Una carpeta con los notebooks o archivos py donde hayas realizado todos los pasos pedidos en el proyecto

### **Consejos.**

- Prioriza cumplir primero con los requisitos mínimos del proyecto de análisis exploratorio de datos (*EDA*), y si lo consideras necesario, agrega pasos adicionales más adelante.
- Asegúrate de que el proyecto te esté ayudando a aprender. Si sientes que no estás adquiriendo nuevos conocimientos, probablemente puedas profundizar más.
  - Dedícale tiempo y trátalo como un caso real: cuida la presentación, realiza análisis significativos y no te quedes en la superficie.
  - ¡Disfruta del proceso y acepta que la frustración es parte del aprendizaje!

### **Criterios de evaluación:**

- **Transformación y limpieza de los datos:** Capacidad para detectar y corregir errores, manejar datos faltantes y realizar modificaciones adecuadas a las columnas y tipos de datos.
- **Uso de los conceptos cubiertos en los módulos de “Python” y “Python for data”:** Demostrar un dominio claro de estructuras de datos como listas, diccionarios, funciones, manejo de archivos, y uso eficiente de *Pandas* para la manipulación de datos.
- **Análisis descriptivo de los datos:** Realizar un análisis estadístico adecuado para describir los principales atributos del conjunto de datos (medias, medianas, desviaciones estándar, correlaciones, etc.).
- **Visualización:** Crear gráficos claros y efectivos utilizando *matplotlib*, *seaborn* u otras librerías, para ilustrar patrones y relaciones relevantes en los datos.
- **Uso eficiente de pandas:** Realizar operaciones como filtrado, agrupamiento, agregaciones, creación de nuevas columnas, y combinaciones de *dataframes* para extraer insights de los datos.
- **Optimización del código en Python:** Aplicar buenas prácticas de programación, como evitar duplicidades, uso eficiente de bucles y comprensión de listas.
- **Informe explicativo del análisis:** Presentar de manera clara los resultados del análisis con justificaciones basadas en datos y conclusiones bien fundamentadas.
- **Readme del proyecto:** Incluir un *README* detallado que describa el propósito del proyecto, los pasos para ejecutarlo y los principales hallazgos.